Mean field simulation for Monte Carlo integration

Part V : Some theoretical analysis

P. Del Moral

INRIA Bordeaux & Inst. Maths. Bordeaux & CMAP Polytechnique

Lectures, INLN CNRS & Nice Sophia Antipolis Univ. 2012

Some hyper-refs

- **Mean field simulation for Monte Carlo integration**. Chapman & Hall - Maths & Stats [600p.] (May 2013).
- **Feynman-Kac formulae**, Genealogical & Interacting Particle Systems with appl., Springer [573p.] (2004)
- **Concentration Inequalities for Mean Field Particle Models**. `Ann. Appl. Probab` (2011). *(+ Rio)*.
- **More references on the website :http://www.math.u-bordeaux1.fr/∼pdelmora/publications.html**

Basic notation

- Dobrushin's contraction coefficient M Markov $E_1 \rightsquigarrow E_2$

$$\begin{aligned}\beta(M) &:= \sup \{\mathrm{osc}(M(f))\ ;\ f \in \mathrm{Osc}(E_2)\} \\ &= \sup\{\|M(x,.) - M(y,.)\|_{\rm tv}\ ;\ (x,y) \in E_1^2\}\end{aligned}$$

- Boltzmann-Gibbs transformation ($G \in (0,1]$)

$$\Psi_G(\mu) = \mu S_{\mu,G}$$

with

$$S_{\mu,G}(x,dy) = G(x)\ \delta_x(dy) + (1 - G(x))\ \Psi_G(\mu)(dy)$$

Properties

$$\Psi_G(\mu) - \Psi_G(\nu) = \frac{1}{\nu(G)}\ (\mu - \nu) S_\mu \quad \text{and} \quad \beta(S_{\mu,G}) \leq 1 - \|G\|$$

Proof:

$$\Psi_G(\mu) - \Psi_G(\nu) = (\mu - \nu) S_\mu + \nu(S_\mu - S_\nu)$$

and

$$\nu(S_\mu - S_\nu) = (1 - \nu(G))\,[\Psi_G(\mu) - \Psi_G(\nu)]$$

Nonlinear semigroups

Normalized & Unnormalized semigroups

$$\Phi_{p,n}(\eta_p) = \eta_n \quad \text{and} \quad \gamma_p Q_{p,n} = \gamma_n$$

Linear integral operators

$$Q_{p,n}(f_n)(x_p) := \mathbb{E}\left(f_n(X_n) \prod_{p \leq q < n} G_q(X_q) \mid X_p = x_p \right)$$

Simplified notation $Q_{n-1,n}(x, dy) = Q_n(x, dy) \, (= G_{n-1}(x) M_n(x, dy))$

$$Q_{p,n} = Q_{p+1} Q_{p+2} \ldots Q_n$$

Nonlinear updating-prediction transformations

$$\Phi_{p,n} = \Phi_n \circ \Phi_{n-1} \circ \ldots \circ \Phi_{p+1}$$

Lipschitz's regularity

$$Q_{p,n}(1)(x) = G_{p,n}(x) \quad \text{and} \quad P_{p,n}(f) = \frac{Q_{p,n}(f)}{Q_{p,n}(1)}$$

$$\Downarrow$$

$$\eta_n(f) = \Phi_{p,n}(\eta_p)(f) = \Psi_{G_{p,n}}(\eta_p)P_{p,n}$$

$$\Downarrow$$

Lipschitz's regularity

$$\left\|\Phi_{p,n}(\eta_p) - \Phi_{p,n}(\eta_p')\right\|_{\rm tv} \leq g_{p,n}\ \beta(P_{p,n})\ \left\|\eta_p - \eta_p'\right\|_{\rm tv}$$

with

$$g_{p,n} := \sup_{x,y} \frac{G_{p,n}(x)}{G_{p,n}(y)}$$

Notation

$$g_n = g_{n,n+1} = \sup_{x,y}(G_n(x)/G_n(y))$$

- **Uniform "contraction" parameter**

$$P_{p,n}(x,dy) = \Phi_{p,n}(\delta_x) \Rightarrow \beta(P_{p,n}) := \sup_{\mu,\nu} \|\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)\|_{tv}$$

- **Uniform potential oscillations**

$$Q_{p,n}(1)(x) = G_{p,n}(x) = \prod_{p \leq q < n} \Phi_{p,q}(\delta_x)(G_q)$$

$$\Downarrow$$

$$\log \frac{G_{p,n}(x)}{G_{p,n}(y)} = \sum_{p \leq q < n} [\log \Phi_{p,q}(\delta_x)(G_q) - \log \Phi_{p,q}(\delta_y)(G_q)]$$

$$\Downarrow \quad \left(\log(y) - \log(x) = \int_0^1 \frac{(y-x)}{x + t(y-x)} dt \right)$$

$$\log g_{p,n} := \sup_{x,y} \log \frac{G_{p,n}(x)}{G_{p,n}(y)} \leq \sum_{p \leq q < n} (g_q - 1)\ \beta(P_{p,q})$$

Contraction properties

Hypothesis

$$\beta(P_{p,n}) \rightarrow_{(n-p)\uparrow\infty} 0 \quad \text{such that} \quad \upsilon := \sup_{p\geq 0} \sum_{p\leq n} (g_n - 1)\, \beta(P_{p,n}) < \infty$$

$$\Downarrow$$

Contraction property : $\exists m \geq 1$

$$\|\Phi_{p,p+m}(\eta) - \Phi_{p,p+m}(\eta')\|_{\rm tv} \leq e^{-1}\, \|\eta - \eta'\|_{\rm tv}$$

Quantitative contraction properties

Key observation

$$P_{p,n}(f) = \frac{M_{p+1}(Q_{p+1,n}(f))}{M_{p+1}(Q_{p+1,n}(1))} = \frac{M_{p+1}(G_{p+1,n}\ P_{p+1,n}(f))}{M_{p+1}(G_{p+1,n})} = R^{(n)}_{p+1} P_{p+1,n}(f)$$

with the triangular array of Markov transitions

$$R^{(n)}_{p+1}(f) := \frac{M_{p+1}(G_{p+1,n} f)}{M_{p+1}(G_{p+1,n})} \Rightarrow P_{p,n} = R^{(n)}_{p+1} R^{(n)}_{p+2} \ldots R^{(n)}_{n}$$

Strong mixing condition $M_n(x, dy) \leq \chi M_n(x', dy)$

$$R^{(n)}_{p+1}(x, dy) := \frac{M_{p+1}(x, dy) G_{p+1,n}(y)}{M_{p+1}(G_{p+1,n})} \leq \chi^2\ R^{(n)}_{p+1}(x', dy)$$

$$\Longrightarrow \beta(R^{(n)}_{p+1}) \leq 1 - \chi^{-2} \Longrightarrow \beta(P_{p,n}) \leq \left(1 - \chi^{-2}\right)^{n-p}$$

Contraction properties

Second key observation : **Mixing condition** $\oplus\ G_n(x) \leq g G_n(y)$

$$\Rightarrow \frac{G_{p,n}(x)}{G_{p,n}(y)} = \frac{Q_{p,n}(1)(x)}{Q_{p,n}(1)(y)} = \frac{G_p(x)}{G_p(y)}\ \frac{M_{p+1}(G_{p+1,n})(x)}{M_{p+1}(G_{p+1,n})(y)} \leq g\ \chi$$

$$\Downarrow$$

Theorem : Stong contraction property

$$\left\|\Phi_{p,n}(\eta_p) - \Phi_{p,n}(\eta'_p)\right\|_{\rm tv} \leq g\ \chi\ \left(1-\chi^{-2}\right)^{n-p}\ \left\|\eta_p - \eta'_p\right\|_{\rm tv}$$

Extensions :
Weak formulation, $M_{p,p+m}(x,dy) \leq \chi_m M_{p,p+m}(x',dy)$, $g\beta(M) < 1$, *etc.*

Orlicz' norm and Gaussian moments

$\pi_\psi[Y]$ Orlicz norm of Y, $\psi(u) = e^{u^2} - 1$

$$\pi_\psi(Y) = \inf\left\{a \in (0,\infty)\ :\ \mathbb{E}(\psi(|Y|/a)) \leq 1\right\}$$

U Gaussian and centered random variable U, s.t. $E(U^2) = 1$:

$$\pi_\psi(U) = \sqrt{8/3}$$

and

$$\begin{aligned} \mathbb{E}\left(U^{2m}\right) &= b(2m)^{2m} := (2m)_m \ 2^{-m} \\ \mathbb{E}\left(|U|^{2m+1}\right) &\leq b(2m+1)^{2m+1} := \frac{(2m+1)_{(m+1)}}{\sqrt{m+1/2}} \ 2^{-(m+1/2)} \end{aligned}$$

Orlicz' norm properties

5 key properties ((Y_i, Y) positive):

1. $Y_1 \leq Y_2 \Longrightarrow \pi_\psi(Y_1) \leq \pi_\psi(Y_2)$
2. $\left(\forall m \geq 0 \quad \mathbb{E}\left(Y_1^{2m}\right) \leq \mathbb{E}\left(Y_2^{2m}\right)\right) \Rightarrow \pi_\psi(Y_1) \leq \pi_\psi(Y_2)$
3. $(\pi_\psi(f(x, Y)) \leq c \quad \text{for } \mathbb{P}\text{-a.e. x}) \Longrightarrow \pi_\psi(f(X, Y)) \leq c$
4. $\mathbb{E}\left(Y^{2m}\right) \leq m! \ \pi_\psi(Y)^{2m}$
5. $\mathbb{E}\left(e^{tY}\right) \leq \min\left(2 \ e^{\frac{1}{4}(t\pi_\psi(Y))^2} \ , \ (1 + t\pi_\psi(Y)) \ e^{(t\pi_\psi(Y))^2}\right)$

$$\Rightarrow \mathbb{P}\left(Y \leq \pi_\psi(Y) \ \sqrt{x + \log 2}\right) \geq 1 - e^{-x}$$

Empirical processes

$$X^i \text{ independent } \sim \mu^i \to m(X) := \frac{1}{N}\sum_{i=1}^{N} \delta_{X^i} \quad \text{and} \quad \mu := \frac{1}{N}\sum_{i=1}^{N} \mu^i$$

Fluctuation centered random fields

$$\begin{aligned} V(X) &= \sqrt{N}\ (m(X) - \mu) \\ \sigma(f)^2 &= \mathbb{E}\left(V(X)(f)^2\right) = \frac{1}{N}\sum_{i=1}^{N} \mu^i([f - \mu^i(f)]^2) \end{aligned}$$

$\mathcal{F}$ separable class of functions $\|f\| \leq 1$

$$\|\mu - \nu\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mu(f) - \nu(f)|,$$

$$\begin{aligned} \mathcal{N}(\epsilon, \mathcal{F}) &= \sup\left\{\mathcal{N}(\epsilon, \mathcal{F}, \mathbb{L}_2(\eta));\ \eta \in \mathcal{P}(E)\right\} \\ I(\mathcal{F}) &= \int_0^2 \sqrt{\log\left(1 + \mathcal{N}(\epsilon, \mathcal{F})\right)}\, d\epsilon \end{aligned}$$

Some useful properties ($G(x) \in [0,1]$, M Markov)

$\rightsquigarrow$ Two classes of functions

$$G \cdot M(\mathcal{F}) = \{G\ M(f)\ :\ f \in \mathcal{F}\}$$
$$G \cdot (M - \mu M)(\mathcal{F}) = \{G\ [M(f) - \mu M(f)]\ :\ f \in \mathcal{F}\}$$

$\Downarrow$ [*Exercice*]

$$\mathcal{N}\left[G \cdot M(\mathcal{F}), \epsilon\right] \leq \mathcal{N}(\mathcal{F}, \epsilon)$$
$$\mathcal{N}\left[G \cdot (M - \mu M)(\mathcal{F}), 2\epsilon\beta(M)\right] \leq \mathcal{N}(\mathcal{F}, \epsilon)$$

Kinchine's inequalities ($\mathrm{osc}(f) \leq 1$)

- Marginal models

$$\mathbb{E}(|V(X)(f)|^m)^{1/m} \leq b(m)\ \mathrm{osc}(f)$$

$$\Downarrow$$

$$\pi_\psi(V(X)(f)) \leq \sqrt{8/3}$$

- Empirical processes

$$\pi_\psi\left(\|V(X)\|_{\mathcal{F}}\right) \leq c\ I(\mathcal{F})$$

Laplace techniques

Legendre-Fenchel transform

$$\forall \lambda \geq 0 \qquad L^{\star}(\lambda) := \sup_{t \in \mathrm{Dom(L)}} (\lambda t - L(t))$$

$L_A(t) := \log \mathbb{E}(e^{tA}) \rightsquigarrow$ Cramér-Chernov-Chebychev inequalities

$$\log \mathbb{P}(A \geq \lambda) \leq -L_A^{\star}(\lambda) \quad \text{and} \quad \mathbb{P}\left(A \geq (L_A^{\star})^{-1}(x)\right) \leq e^{-x}$$

- Comparison property

$$L_1 \leq L_2 \Rightarrow L_2^{\star} \leq L_1^{\star} \Rightarrow (L_1^{\star})^{-1} \leq (L_2^{\star})^{-1}$$

- J. Bretagnolle & E. Rio's Lemma

$$(L_{A+B}^{\star})^{-1}(x) \leq (L_A^{\star})^{-1}(x) + (L_B^{\star})^{-1}(x)$$

3 examples-exercices

- $L(t) = t^2/(1-t),\ t \in [0,1[$

$$L^\star(\lambda) = \left(\sqrt{\lambda+1} - 1\right)^2 \quad \& \quad (L^\star)^{-1}(x) = \left(1+\sqrt{x}\right)^2 - 1 = x + 2\sqrt{x}$$

- $L_0(t) := -t - \frac{1}{2}\log(1-2t),\ t \in [0,1/2[$

$$L_0^\star(\lambda) = \frac{1}{2}(\lambda - \log(1+\lambda)) \quad \& \quad (L_0^\star)^{-1}(x) \le 2(x+\sqrt{x})$$

- $L_1(t) := e^t - 1 - t$

$$L_1^\star(\lambda) = (1+\lambda)\log(1+\lambda) - \lambda \quad \& \quad (L_1^\star)^{-1}(x) \le \frac{x}{3} + \sqrt{2x}$$

Applications (part 1)

- Centered $A \leq 1$ & $\sigma_A = \mathbb{E}(A^2)^{1/2} \Rightarrow L_A(t) \leq \sigma_A^2\ L_1(t)$
  $\Rightarrow$ The probability of the following events is greater than $1 - e^{-x}$

$$A \leq \sigma_A^2\ (L_1^\star)^{-1}\left(\frac{x}{\sigma_A^2}\right) \leq \frac{x}{3} + \sigma_A\ \sqrt{2x}$$

- B s.t. $\mathbb{E}\left(|B|^m\right)^{1/m} \leq b(2m)^2\ c \Rightarrow L_B(t) \leq ct + L_0(ct)$
  $\Rightarrow$ The probability of the following events

$$B \leq c\left[1 + (L_0^\star)^{-1}(x)\right] \leq c\left[1 + 2(x + \sqrt{x})\right]$$

  is greater than $1 - e^{-x}$.

- Concentration of $A + B$ using J. Bretagnolle & E. Rio's Lemma

Applications (part 2)

- $0 < \mathrm{osc}(\mathrm{f}) \leq a \Rightarrow L_{\sqrt{N}V(X)(f)}(t) \leq N\ \sigma^2(f/a)\ L_1(at)$

  $\Rightarrow$ the probability of the following events is greater than $1 - e^{-x}$,

$$V(X)(f) \quad \leq \quad a^{-1}\sigma^2(f)\sqrt{N}\ \ (L_1^\star)^{-1}\left(\frac{xa^2}{N\sigma^2(f)}\right) \leq \frac{xa}{3\sqrt{N}} + \sqrt{2x\sigma(f)^2}$$

- Concentration of $F(m(X)(f))$ [marginal or empirical processes] using J. Bretagnolle & E. Rio's Lemma

Interacting processes

Markov $X_n = (X_n^i)_{1\leq i\leq N} \in E_n^N$, conditionally independent $\mid \mathcal{G}_{n-1}$.

$$\mu_n^i = \mathrm{Law}(X_n^i \mid X_0, \ldots, X_{n-1}) \rightsquigarrow V(X_n) := \sqrt{N}\,(m(X_n) - \mu_n)$$

Definition.: $f_n \in \mathcal{G}_{n-1}$, $\mathrm{osc}(f_n) \leq 1 \rightsquigarrow \mathbb{E}\left(V(X_n)(f_n)^2 \,|\mathcal{G}_{n-1}\right) \leq \sigma_n^2$

$$\overline{\sigma}_n^2 := \sum_{0\leq p\leq n} \sigma_p^2 \quad \text{and} \quad a_n^\star := \max_{0\leq p\leq n} a_p$$

- $V_n(X)(f) = \sum_{p=0}^n a_p \; V(X_p)(f_p)$

$$L_{\sqrt{N}V_n(X)(f)}(t) \leq N\;\overline{\sigma}_n^2\; L_1(ta_n^\star)$$

$\Rightarrow$ the probability of the following events is greater than $1 - e^{-x}$

$$V_n(X)(f) \quad \leq \quad \sqrt{N}\; a_n^\star\; \overline{\sigma}_n^2\; \left(L_1^\star\right)^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right) \leq a_n^\star \left(\frac{x}{3\sqrt{N}} + \sqrt{2\overline{\sigma}_n^2\; x}\right)$$

Perturbation analysis (marginal models)

$$W_n(X)(f) = V_n(X)(f) + \frac{1}{\sqrt{N}}\ R_n(X)(f)$$

with

$$V_n(X)(f) = \sum_{p=0}^{n} a_p\ V(X_p)(f_p) \quad \& \quad \mathbb{E}\left(|R_n(X)(f)|^m\right)^{1/m} \leq b(2m)^2\ r_n$$

Using J. Bretagnolle & E. Rio's Lemma

$\Rightarrow$ the probability of the following events is greater than $1 - e^{-x}$,

$$\sqrt{N}\ W_n(X)(f) \leq r_n\ \left(1 + (L_0^\star)^{-1}(x)\right) + N\ a_n^\star\ \overline{\sigma}_n^2\ (L_1^\star)^{-1}\left(\frac{x}{N\overline{\sigma}_n^2}\right)$$

Perturbation analysis (empirical processes)

$$W_n(X)(f) = V_n(X)(f) + \frac{1}{\sqrt{N}}\ R_n(X)(f)$$

with

$$V_n(X)(f) = \sum_{p=0}^{n} a_p\ V(X_p)(f_p) \quad \& \quad \mathbb{E}\left(\|R_n(X)\|_{\mathcal{F}}^m\right) \leq m!\ r_n^m$$

Using J. Bretagnolle & E. Rio's Lemma
$\Rightarrow$ the probability of the following events is greater than $1 - e^{-x}$,

$$\|W_n(X)\|_{\mathcal{F}} \leq c\ \left[\sum_{p=0}^{n} a_p\right]\ I(\mathcal{F})\left(1+2\sqrt{x}\right) + \frac{r_n}{\sqrt{N}}\left(1+(L_0^{\star})^{-1}\left(\frac{x}{2}\right)\right)$$

First order expansions

Key telescoping decomposition

$$\eta_n^N - \eta_n = \sum_{p=0}^{n} \left[\Phi_{p,n}(\eta_p^N) - \Phi_{p,n}\left(\Phi_p(\eta_{p-1}^N)\right)\right]$$

$\oplus$ First order expansion

$$\sqrt{N}\Phi_{p,n}(\eta_p^N) - \Phi_{p,n}\left(\Phi_p(\eta_{p-1}^N)\right)$$

$$= \sqrt{N}\Phi_{p,n}\left(\Phi_p(\eta_{p-1}^N) + \frac{1}{\sqrt{N}} \; V_p^N\right) - \Phi_{p,n}\left(\Phi_p(\eta_{p-1}^N)\right)$$

$$\simeq V_p^N D_{p,n} + \tfrac{1}{\sqrt{N}} \; R_{p,n}^N$$

with $D_{p,n} \in \mathcal{G}_{p-1}$-first order integral operator $\oplus$ 2nd-order remainder $R_{p,n}^N$

First order expansions

Stochastic perturbation model

$$W_n^{\eta,N} := \sqrt{N}\left[\eta_n^N - \eta_n\right] = \sum_{0\leq p\leq n} V_p^N D_{p,n} + \frac{1}{\sqrt{N}}\ R_n^N$$

Under the mixing condition of FK semigroups

$$\mathrm{osc}\left(D_{p,n}(f)\right) \leq c\ g_{p,n}\ \beta(P_{p,n}) \leq c(1-\epsilon)^{n-p}$$

and

$$\mathbb{E}\left(|R_n^N(f)|^m\right) \leq b(2m)^{2m} c$$

$$\Downarrow$$

Uniform concentrtation estimates w.r.t. the time parameter

Particle free energy

Multiplicative formulae

$$\mathcal{Z}_n^N = \prod_{0\leq p<n} \eta_p^N(G_p) = \gamma_n^N(1) \longrightarrow_{N\uparrow\infty} \gamma_n(1) = \prod_{0\leq p<n} \eta_p(G_p)$$

Taylor first order expansion

$$\forall x, y > 0 \qquad \log y - \log x = \int_0^1 \frac{(y-x)}{x+t(y-x)}\ dt$$

$$\Downarrow$$

$$\begin{aligned}
&\log\left(\gamma_n^N(1)/\gamma_n(1)\right) \\
&= \sum_{0\leq p<n} \left(\log \eta_p^N(G_p) - \log \eta_p(G_p)\right) \\
&= \sum_{0\leq p<n} \left(\log\left(\eta_p(G_p) + \frac{1}{\sqrt{N}}\ W_p^{\eta,N}(G_p)\right) - \log \eta_p(G_p)\right) \\
&= \frac{1}{\sqrt{N}} \sum_{0\leq p<n} \int_0^1 \frac{W_p^{\eta,N}(G_p)}{\eta_p(G_p) + \frac{t}{\sqrt{N}}\ W_p^{\eta,N}(G_p)}\ dt
\end{aligned}$$

$\rightsquigarrow$ first order expansion [exo]